AD-A250 144

:NTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Information and Regulatory Affairs, Office of Management and Budget, Washington, DC 20503.

1. AGENCY USE ONLY *(Leave Blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	1986	Unknown

4. TITLE AND SUBTITLE	5. FUNDING NUMBERS
Epistemological Relevance and Statistical Knowledge	DAAB10-86-C-0567

6. AUTHOR(S)

Henry E. Kyburg, Jr.

DTIC
ELECTE
MAY 13 1992
S D

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
University of Rochester Department of Philosophy Rochester, NY 14627	

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSORING/MONITORING AGENCY REPORT NUMBER
U.S. Army CECOM Signals Warfare Directorate Vint Hill Farms Station Warrenton, VA 22186-5100	92-TRF-0002

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT	12b. DISTRIBUTION CODE
Statement A; Approved for public release; distribution unlimited.	

13. ABSTRACT *(Maximum 200 words)*

For many years, at least since McCarthy and Hayes (1969), writers have lamented, and attempted to compensate statistical knowledge for governing the uncertainty of belief, for making uncertain inference, and the like. It is hardly ever spelled out what "adequate statistical knowledge" would be, if we had it, and how adequate statistical knowledge could be used to control and regulate epistemic uncertainty. One response to lack of adequate statistics has been to search for non-statistical measures of uncertainty. The minimal variant has been to propose "subjective probability" as a concept to which we can turn when we lack statistics.

14. SUBJECT TERMS		15. NUMBER OF PAGES
Artificial Intelligence, Data Fusion, Espistemological Relevance Statistical Knowledge		5
		16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Epistemological Relevance and Statistical Knowledge*

by

Henry E. Kyburg, Jr.
University of Rochester

topic: Uncertainty Propagation

QUALITY INSPECTED 4

92-12585

Epistemological Relevance and Statistical Knowledge

We addresss the question of how far you can go in propagating uncertainty according to statistical knowledge, as opposed to depending on subjective opinion.

1. Background.

For many years, at least since McCarthy and Hayes (1969), writers have lamented, and attempted to compensate for, the alleged fact that we often do not have adequate statistical knowledge for governing the uncertainty of belief, for making uncertain inferences, and the like. It is hardly ever spelled out what "adequate statistical knowledge" would be, if we had it, and how adequate statistical knowledge could be used to control and regulate epistemic uncertainty.

One response to the lack of adequate statistics has been to search for non-statistical measures of uncertainty. The minimal variant has been to propose "subjective probability" as a concept to which we can turn when we lack statistics.

2. Assumptions.

The assumptions we make here are relatively few. We suppose that the knowledge base may have general statistical knowledge in it.

Our second assumption is that statements fall into equivalence classes with respect to the statistical information that is epistemically most relevant to them. We can express this as a formal principle: If "$S \rightleftarrows T$" is in our knowledge base, then the same statistical knowledge is potentially relevant to S and to T.

Our third assumption is the general one that our knowledge base can be expressed in a first order extensional language. We take an individual, however, to be arbitrarily complex: for example it might be a trial of a complicated compound experiment.

Finally, in order for statistics to be of interest, we suppose that we may know some things about an individual without knowing everything about it.

3. Interference I.

We will be concerned with the way in which some items of statistical knowledge can interfere with the epistemic relevance

of other items. One principle bearing on this is the subset
principle:

<u>The Subset Principle</u>: Suppose that "a is a B" is in our knowledge
base, and that "$\%(B, C) = p$" is in our knowledge base. Suppose
that we know that a' is a C' if and only if a is a C, that a' is
a B', and that $\%(B', C') = p'$, where $p \neq p'$. This statistical
knowledge is *epistemically irrelevant* if we know of a subset of
B', B'', such that we know both a' is a B'' and $\%(B'', C') =$
p.

The subset principle is one that has been frequently
identified in the context of non-monotonic logic.

4. Interference II.
Some situations call for a second principle. Suppose we
have a roomful of urns, and that a designates a ball in the
room. Suppose we know that there are 100 balls in the room,
and that 50 are black. But suppose we also know that there are
10 urns, that 9 of them containing four black balls and one
white ball, and that the tenth contains the remainder of the
balls. The relative frequency of black balls in the first nine urns
is .8, and the relative frequency of black balls in the tenth urn
is $14/45 = .311$.
What statistics are relevant to the statement, "a is
black." If we know of a only that it is a ball in the room, it is
only the statistics about the frequency of black balls in the room
that are relevant. If we know <u>also</u> something about how a came
to be the designated ball, the other statistics may also be
relevant. For example, we might know that a is the ball
resulting from first choosing an urn at random, and then
choosing a ball at random from the chosen urn. If that is the
case, the relevant statistics are those governing the proportion of
pairs consisting of an urn, and a ball drawn from that urn, such
that the second member of the pair is black. We call the rule
governing this situation the Bayesian Principle:

<u>The Bayesian Principle</u>: Suppose that "$<a, b>$ is a B" is in our
knowledge base, and that "$\%(B, C) = p$" is in our knowledge
base. Suppose that we know that a' is a C' if and only if a is
a C, that a' is a B', and that $\%(B', C') = p' \neq p$. This

2

statistical knowledge is *epistemically irrelevant* if we know of a cross product of B' with B'' and a corresponding subset C'' and a'' such that

     (1) $<a', a''>$ is known to be in $B' \times B''$,
     (2) $<a', a''>$ is in C'' if and only if a is in C,
     (3) $\%(B' \times B'', C'') = p'$,

and for some B^* known to be a subset of $B' \times B''$,

     (4) $\%(B^*, C) = p$.

5. Interference III.

The final principle of relevance we need for dealing with statistical knowledge is in a sense the dual of our first principle, the subset principle.

<u>The Supersample Principle</u>: Suppose that we know that a_n is a member of P^n and that we are interested in the chance that a_n is Q (e.g., "representative within ϵ"). Suppose that a is known to be a member of R, that a is a Q' if and only if a_n is Q, and that $\%(P^n, Q) = p \neq p' = \%(R, Q')$ are all known. Then our statistical knowledge about R is *epistemically irrelevant* if there is a parallel structure to our original one that is such that we also know that a is a subset of a_n.

6. Conclusions:

We arrive at several conclusions.

(1) If we accept the equivalence condition -- that statements connnected in our knowledge base by a biconditional should have the same probability -- then many more statements than might at first have been thought can have probabilities based on statistical background knowledge

(2) This has a profound bearing on the representation of uncertainty in our bodies of knowledge. If we suppose that "subjective" confidence in the strong sense of "subjective" is acceptable as a measure of uncertainty only when statistical information is not available, then there are far fewer situations in which purely subjective uncertainties are called for than some people have suggested

(3) Given the equivalence condition, there may be many potential reference sets for a given equivalence class of statements. We therefore need a way of adjudicating our choice among these reference sets.

(4) There are three ways in which conflict between two potential reference classes can be resolved to the benefit of one of them. Only one of these ways seems to have worked its way into the literature on non-monotonic logic. All three should be taken account of.

(5) These three resolutions reflect the three principles: the Subset Principle, the Bayesian Principle, and the Superset Principle. (In fact the subset principle is reducible to the Bayesian principle (see Kyburg 1983).)

(6) The results of this analysis can be used to implement probabilistic non-monotonic acceptance as well as to determine rationally allowable distributions of uncertainty.

Henry E. Kyburg, Jr.   University of Rochester

4